End Course Summative Assignment

Problem Statement: Write the Solutions to the Top 50 Interview Questions and Explain any 5 Questions in a Video

Imagine you are a dedicated student aspiring to excel in job interviews. Your task is to write the solutions for any 50 interview questions out of 80 total questions presented to you. Additionally, create an engaging video where you thoroughly explain the answers to any five of these questions.

Your solutions should be concise, well-structured, and effective in showcasing your problem-solving skills. In the video, use a dynamic approach to clarify the chosen questions, ensuring your explanations are easily comprehensible for a broad audience.

Note:

1. Make a copy of this document and write your answers.
2. Include the Video Link here in your document before submitting.

1. What is a vector in mathematics?

Sol:- vector is a quantity which has both magnitude and direction. A vector quantity, unlike scalar, has a direction component along with the magnitude which helps to determine the position of one point relative to the other.

2. How is a vector different from a scalar?

Sol:- scalar quantity is different from a vector quantity in terms of direction. Scalars don't have direction, whereas a vector has. Due to this feature, the scalar quantity can be said to be represented in one dimension, whereas a vector quantity can be multi-dimensional. From the table given below, let us learn more differences between scalars and vectors.

3. What are the different operations that can be performed on vectors?

Sol:- **Vector Operations**

The following include the list of vector operations:

- Addition and Subtraction of vectors
- Multiplication of vectors

4. How can vectors be multiplied by a scalar?

Sol:- When a vector is multiplied by a scalar quantity, the magnitude of the vector changes in proportion to the scalar magnitude, but the direction of the vector stays the same.

5. What is the magnitude of a vector?

Sol:- As we know, the vector is an object which has both the magnitude as well as direction. To find the magnitude of a vector, we need to calculate the length of the

vector. Quantities such as velocity, displacement, force, momentum, etc. are vector quantities. But speed, mass, distance, volume, temperature, etc. are scalar quantities. The scalar has the only magnitude, whereas the vectors have both magnitude and direction.

6. How can the direction of a vector be determined?

Sol:- Finding the direction of a vector in a 2-dimensional plane is easy! You'll just need a little trigonometry. The x and y components of a vector form a right triangle. You can use the tangent function to find the angle between the x-axis and the vector. This wikiHow guide will show you how to find the direction of a vector and walk through four examples. Additionally, we'll review how to find the magnitude of a vector.

7. What is the difference between a square matrix and a rectangular matrix?

Sol:- Matrix is one of the most commonly used elements in linear algebra. Matrix is the rectangular arrangement of numbers/elements/objects. The horizontal arrangement is called the row and the vertical arrangement is the column of a matrix. The order of a matrix is determined by the number of rows by columns. Suppose a matrix has 2 rows and 3 rows of elements, then its order is 2×3. In the same way, when a matrix has an equal number of rows and columns, then the matrix is called the **square matrix**. In this article, you will learn the mathematical definition of a square matrix, properties of square matrix and examples in detail.

8. What is a basis in linear algebra?

Sol:- **basis** for a vector space is a sequence of vectors that form a set that is linearly independent and that spans the space.

9. What is a linear transformation in linear algebra?

Sol:- **linear transformation** is a function from one vector space to another that respects the underlying (linear) structure of each vector space. A linear transformation is also known as a linear operator or map. The range of the transformation may be the same as the domain, and when that happens, the transformation is known as an endomorphism or, if invertible, an automorphism. The two vector spaces must have the same underlying field.

10. What is an eigenvector in linear algebra?

Sol:- Eigenvectors are the vectors (non-zero) that do not change the direction when any linear transformation is applied. It changes by only a scalar factor. In a brief, we can say, if A is a linear transformation from a vector space V and **x** is a vector in V, which is not a zero vector, then v is an eigenvector of A if A(X) is a scalar multiple of **x**.

11. What is the gradient in machine learning?

Sol:- gradient simply measures the change in all weights with regard to the change in error. You can also think of a gradient as the slope of a function. The higher the gradient, the steeper the slope and the faster a model can learn. But if the slope is zero, the model stops learning.

12. What is backpropagation in machine learning?

Sol:- **Backpropagation** is the essence of neural network training. It is the method of fine-tuning the weights of a neural network based on the error rate obtained in the previous epoch (i.e., iteration). Proper tuning of the weights allows you to reduce error rates and make the model reliable by increasing its generalization.

13. What is the concept of a derivative in calculus?

Sol:- **Derivatives** are defined as the varying rate of change of a function with respect to an independent variable. The derivative is primarily used when there is some varying quantity, and the rate of change is not constant.

14. How are partial derivatives used in machine learning?

Sol:- The partial derivative answers the questions of how f changes (∂f) when **one** variable changes by a small amount (∂x). In this setting all other variables are assumed to be constant and static. Thus the partial derivative is denoted $\partial f/\partial x$.

15. What is probability theory?

Sol:- Probability is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty. We can predict only the chance of an event to occur i.e., how likely they are going to happen, using it. Probability can range from 0 to 1, where 0 means the event to be an impossible one and 1 indicates a certain event. Probability for Class 10 is an important topic for the students which explains all the basic concepts of this topic

16. What are the primary components of probability theory?

Sol:- Uncertainty,

Probability,

strength of evidence

17. What is conditional probability, and how is it calculated?

Sol:- **Conditional probability** is known as the possibility of an event or outcome happening, based on the existence of a previous event or outcome. It is calculated by multiplying the probability of the preceding event by the renewed probability of the succeeding, or conditional, event.

18. What is Bayes theorem, and how is it used.

Sol:- Applications of Bayes' Theorem are widespread and not limited to the financial realm. For example, Bayes' theorem can be used to determine the accuracy of medical test results by taking into consideration how likely any given person is to have a disease and the general accuracy of the test

19. What is a random variable, and how is it different from a regular variable?

Sol:- random variable is a function that assigns a number depending upon the outcome of a probabilistic experiment.

For example, a random variable might be the revenue earned for one days operation of a fast food restaurant.

20. What is the law of large numbers, and how does it relate to probability theory?

Sol:- The law of large numbers, in probability and statistics, states that as a sample size grows, its mean gets closer to the average of the whole population. This is due to the sample being more representative of the population as the sample become larger.

21. What is the central limit theorem, and how is it used?

Sol:- The CLT is a statistical theory that states that - if you take a sufficiently large sample size from a population with a finite level of variance, the mean of all from that population will be roughly equal to the population mean.

22. What is the difference between discrete and continuous probability distributions?

Sol:- discrete distribution is one in which the data can only take on certain values, for example integers. A continuous distribution is one in which data can take on any value within a specified range (which may be infinite)

23. What are some common measures of central tendency, and how are they calculated?

Sol:- The central tendency of the dataset can be found out using the three important measures namely mean, mode median.

Mean

The mean represents the average value of the dataset. It can be calculated as the sum of all the values in the dataset divided by the number of values. In general, it is considered as the arithmetic mean. Some other measures of mean used to find the central tendency are as follows:

- Geometric Mean
- Harmonic Mean

- Weighted Mean

Median

Median is the middle value of the dataset in which the dataset is arranged in the ascending order or in descending order. When the dataset contains an even number of values, then the median value of the dataset can be found by taking the mean of the middle two values.

Mode

The mode represents the frequently occurring value in the dataset. Sometimes the dataset may contain multiple modes it does not contain any mode at all.

Consider the given dataset 5, 4, 2, 3, 2, 1, 5, 4, 5

24. What is the purpose of using percentiles and quartiles in data summarization?

Sol:- **Percentiles**

A percentile is a measure at which that percentage of the total values are the same as or below that measure. For example, 90% of the data values lie below the 90th percentile, whereas 10% of the data values lie below the 10th percentile.

Quartiles

Quartiles are values that divide a (part of a) data table into four groups containing an approximately equal number of observations. The total of 100% is split into four equal parts: 25%, 50%, 75% and 100%.

The *first quartile* (or lower quartile), Q1, is defined as the value that has an f-value equal to 0.25. This is the same thing as the twenty-fifth percentile. The *third*

quartile (or upper quartile), Q3, has an f-value equal to 0.75. The interquartile range, IQR, is defined as Q3-Q1.

25. How do you detect and treat outliers in a dataset?

Sol:- When the data, or certain features in the dataset, follow a normal distribution, you can use the standard deviation of the data, or the equivalent z-score to detect outliers.

In statistics, standard deviation measures the *spread of data around the mean*, and in essence, it captures how far away from the mean the data points are.

26. How do you use the central limit theorem to approximate a discrete probability distribution?

Sol:- It is important for you to understand when to use the **central limit theorem**. If you are being asked to find the probability of the mean, use the clt for the means. If you are being asked to find the probability of a sum or total, use the clt for sums. This also applies to percentiles for means and sums.

27. How do you test the goodness of fit of a discrete probability distribution?

Sol:- Discrete probability distributions are based on discrete variables, which have a finite or countable number of values. In this post, I show you how to perform goodness-of-fit tests to determine how well your data fit various discrete probability distributions.

28. What is a joint probability distribution?

Sol:- Probability is a branch of mathematics which deals with the occurrence of a random event. In simple words it is the likelihood of a certain event. A statistical

measure that calculates the likelihood of two events occurring together and at the same point in time is called Joint probability.

29. How do you calculate the joint probability distribution?

Sol:- Notation to represent the joint probability can take a few different forms. The following formula represents the joint probability of events with intersection.

P (A∩B)

where,

A, B= Two events

P(A and B),P(AB)=The joint probability of A and B

The symbol "∩" in a joint probability is called an intersection. The probability of event A and event B happening is the same thing as the point where A and B intersect. Hence, the joint probability is also called the intersection of two or more events. We can represent this relation using a Venn diagram as shown below.

30. What is the difference between a joint probability distribution and a marginal probability distribution?

Sol:- joint distribution is a probability distribution that describes the probability of two or more random variables having specific values at the same time. The joint distribution is represented by a probability density function (pdf) or probability

mass function (pmf) that assigns a probability to each possible combination of values of the random variables. For example, the joint distribution of two random variables X and Y is represented by the function $P(X = x, Y = y)$, where x and y are specific values of the random variables X and Y.

A marginal distribution is a probability distribution that describes the probability of one random variable having a specific value, regardless of the value of any other random variables. The marginal distribution is obtained by summing or integrating the joint distribution over the values of the other random variables. For example, the marginal distribution of a random variable X is represented by the function $P(X = x) = \sum_y P(X = x, Y = y)$, where x is a specific value of the random variable X and y is a variable representing the values of another random variable Y.

31. What is the covariance of a joint probability distribution?

Sol:- Covariance between variables can be calculated in two ways. One method is the historical sample covariance between two random variables Xi and Yi. It is based on a sample of past data of size n and is given by:

$$Cov Xi, Yi = \sum_{ni=1} (Xi - \bar{X})(Yi - \bar{Y}) n - 1$$

Alternatively, covariance can be defined as probability-weighted average of the cross-products of each random variable's deviation from its own expected value. That is:

$$CovXi,Yi=E[(Xi-\bar{}X)(Yi-\bar{}Y)]$$

32. How do you determine if two random variables are independent based on their joint probability distribution?

Sol:- we usually need to deal with more than one random variable. For example, if you study physical characteristics of people in a certain area, you might pick a person at random and then look at his/her weight, height, etc. The weight of the randomly chosen person is one random variable, while his/her height is another one. Not only do we need to study each random variable separately, but also we need to consider if there is *dependence* (i.e., correlation) between them. Is it true that a taller person is more likely to be heavier or not? The issues of dependence between several random variables will be studied in detail later on, but here we would like to talk about a special scenario where two random variables are *independent*.

The concept of independent random variables is very similar to independent events. Remember, two events A and B are independent

33. What is the relationship between the correlation coefficient and the covariance of a joint probability distribution?

Sol:- Covariance measures the degree to which two random variables X and Y are linearly related. It is defined as the expected value of the product of the deviations of X and Y from their respective means:

$$Cov(X<Y) = E[(X-E[X])(Y-E[Y])]$$

Correlation is a standardized version of covariance, and measures the degree of linear association between two variables X and Y:

$$Corr(X<Y) = cov(X,Y)/(std(X)*std(Y))$$

34. What is sampling in statistics, and why is it important?

Sol:- Sampling in statistics and data analytics is the practice of selecting a subset, or sample, of data from a larger population or dataset. This subset is chosen with the aim of making inferences, drawing conclusions, or performing analysis on the entire population based on the characteristics and patterns observed in the sample. Sampling is a crucial technique for several reasons:

35. What are the different sampling methods commonly used in statistical inference?

Sol:- When you conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, you select a **sample**. The sample is the group of individuals who will actually participate in the research.

To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. This is called a **sampling method**. There are two primary types of sampling methods that you can use in your research:

- **Probability sampling** involves random selection, allowing you to make strong statistical inferences about the whole group.
- **Non-probability sampling** involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

36. What is the central limit theorem, and why is it important in statistical inference?

Sol:- The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution.

Sample sizes equal to or greater than 30 are often considered sufficient for the CLT to hold.

A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.

37. What is the difference between parameter estimation and hypothesis testing?

Sol:- Point Estimation deals with the method of estimating an unknown parameter of a population based on Random Samples from the same population. The assumption here is that the parameter to be estimated is a constant with one value and the sample Statistic computed from the sample is estimating that value exactly. In the parameter space, it is represented as a point. Hence the name point estimation. Maximum Likelihood is such a method.

The Testing of Hypothesis is the process of either rejecting or not rejecting a statement or a Hypothesis that has been set up about the Parameter. This is done by computing a Test Statistic based on the sample(s) from that population and testing it against an ideal standard value. A " risk" of a making wrong conclusion is predetermined. That means the probability of rejecting the original hypothesis when it is a correct one. The basic hypothesis about the parameter is known as the Null hypothesis and an Alternative hypothesis is also set up at the same time. In case the Null (H0) is rejected, the Alternative (H1) is accepted.

38. What is the p-value in hypothesis testing?

Sol:- The P value is defined as the probability under the assumption of no effect or no difference (null hypothesis), of obtaining a result equal to or more extreme than what was actually observed. The P stands for probability and measures how likely it is that any observed difference between groups is due to chance.

38. What is confidence interval estimation?

Sol:- A confidence interval is the mean of your estimate plus and minus the variation in that estimate. This is the range of values you expect your estimate to fall between if you redo your test, within a certain level of confidence.

39. What are Type I and Type II errors in hypothesis testing?

Sol:- A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

40. What is the difference between correlation and causation?

Sol:- **Correlation** describes an association between types of variables: when one variable changes, so does the other. A correlation is a statistical indicator of the relationship between variables. These variables change together: they covary. But this covariation isn't necessarily due to a direct or indirect causal link.

Sol:- means that changes in one variable brings about changes in the other; there is a cause-and-effect relationship between variables. The two variables are correlated with each other and there is also a causal link between them.

41. How is a confidence interval defined in statistics?

Sol:- A confidence interval, in statistics, refers to the probability that a population parameter will fall between a set of values for a certain proportion of times. Analysts often use confidence intervals that contain either 95% or 99% of expected observations. Thus, if a point estimate is generated from a statistical model of

10.00 with a 95% confidence interval of 9.50 - 10.50, it means we are 95% confident that the true value falls within that range.

42.How is a confidence interval defined in statistics?

Sol:- In Statistics, a **confidence interval** is a kind of interval calculation, obtained from the observed data that holds the actual value of the unknown parameter. It is associated with the confidence level that quantifies the confidence level in which the interval estimates the deterministic parameter. Also, we can say, it is based on Standard Normal Distribution, where Z value is the z-score. Here, let us look at the definition, formula, table, and the calculation of the confidence level in detail.

43. What does the confidence level represent in a confidence interval?

Sol:- Confidence levels and confidence intervals are measures within statistics that show the certainty of a sample being a good representation of the whole population. Though they can work together, the two measures have many differences including definition, calculation method and expression. Learning more about confidence levels and confidence intervals can help you better understand samples and their relation to the population. In this article, we explain what confidence levels are, define confidence intervals, list the differences between the two and offer tips for learning to use statistics.

44. What is hypothesis testing in statistics?

Sol:- Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test. It is used to estimate the relationship between 2 statistical variables.

45. What is the purpose of a null hypothesis in hypothesis testing?

Sol:- The **null hypothesis** is a kind of hypothesis which explains the population parameter whose purpose is to test the validity of the given experimental data. This hypothesis is either rejected or not rejected based on the viability of the given population or sample.

46. What is the difference between a one-tailed and a two-tailed test?

Sol:- The two ways of carrying out statistical significance test of a characteristic, drawn from the population, with respect to the test statistic, are a one-tailed test and two-tailed test. The **one-tailed test** refers to a test of null hypothesis, in which the alternative hypothesis is articulated directionally. Here, the critical region lies only on one tail. However, if the alternative hypothesis is not exhibited directionally, then it is known as the **two-tailed test** of the null hypothesis., wherein the critical region is one both the tails.

47. What is experiment design, and why is it important?

Sol:- Experimental design is the process of carrying out research in an objective and controlled fashion so that precision is maximized and specific conclusions can be drawn regarding a hypothesis statement. Generally, the purpose is to establish the effect that a factor or independent variable has on a dependent variable.

48. What are the key elements to consider when designing an experiment?

Sol:- **Key things to consider when designing your experiment:**

Ensure your experiment is unbiased

Make sure your experiment is adequately powered

Consider the range of applicability of your experiment

Simplify your experiment

Indicate the uncertainty in your results

References and further reading

49. How can sample size determination affect experiment design?

Sol:- Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample. In practice, the sample size used in a study is usually determined based on the cost, time, or convenience of collecting the data, and the need for it to offer sufficient statistical power. In complicated studies there may be several different sample sizes: for example, in a stratified survey there would be different sizes for each stratum. In a census, data is sought for an entire population, hence the intended sample size is equal to the population. In experimental design, where a study may be divided into different treatment groups, there may be different sample sizes for each group.

50. What are some strategies to mitigate potential sources of bias in experiment design?

Sol:- **Research bias** results from any deviation from the truth, causing distorted results and wrong conclusions. Bias can occur at any phase of your research, including during data collection, data analysis, interpretation, or publication. Research bias can occur in both qualitative and quantitative research.

Understanding research bias is important for several reasons.

1. Bias exists in all research, across research designs, and is difficult to eliminate.
2. Bias can occur at any stage of the research process.
3. Bias impacts the validity and reliability of your findings, leading to misinterpretation of data.